The complete genome sequence of *Mycobacterium bovis*

Thierry Garnier*, Karin Eiglmeier*, Jean-Christophe Camus*†, Nadine Medina*, Huma Mansoor‡, Melinda Pryor*†, Stephanie Duthoy*, Sophie Grondin*, Celine Lacroix*, Christel Monsempe*, Sylvie Simon*, Barbara Harris§, Rebecca Atkin§, Jon Doggett§, Rebecca Mayes§, Lisa Keating‡, Paul R. Wheeler‡, Julian Parkhill§, Bart G. Barrell§, Stewart T. Cole*, Stephen V. Gordon‡¶, and R. Glyn Hewinson‡

*Unité de Génétique Moléculaire Bactérienne and †PT4 Annotation, Génopole, Institut Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France; ‡Tuberculosis Research Group, Veterinary Laboratories Agency Weybridge, Woodham Lane, New Haw, Addlestone, Surrey KT15 3NB, United Kingdom; and §The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, United Kingdom

Edited by John J. Mekalanos, Harvard Medical School, Boston, MA, and approved March 19, 2003 (received for review January 24, 2003)

***Mycobacterium bovis* is the causative agent of tuberculosis in a range of animal species and man, with worldwide annual losses to agriculture of $3 billion. The human burden of tuberculosis caused by the bovine tubercle bacillus is still largely unknown. *M. bovis* was also the progenitor for the *M. bovis* bacillus Calmette–Guérin vaccine strain, the most widely used human vaccine. Here we describe the 4,345,492-bp genome sequence of *M. bovis* AF2122/97 and its comparison with the genomes of *Mycobacterium tuberculosis* and *Mycobacterium leprae*. Strikingly, the genome sequence of *M. bovis* is >99.95% identical to that of *M. tuberculosis*, but deletion of genetic information has led to a reduced genome size. Comparison with *M. leprae* reveals a number of common gene losses, suggesting the removal of functional redundancy. Cell wall components and secreted proteins show the greatest variation, indicating their potential role in host–bacillus interactions or immune evasion. Furthermore, there are no genes unique to *M. bovis*, implying that differential gene expression may be the key to the host tropisms of human and bovine bacilli. The genome sequence therefore offers major insight on the evolution, host preference, and pathobiology of *M. bovis*.**

In his Nobel Prize address of 1901 Von Behring stated, "As you know, tuberculosis in cattle is one of the most damaging infectious diseases to affect agriculture" (www.nobel.se/medicine/laureates/1901). The past 100 years of research has had little impact on this conclusion in developing countries, whereas in some countries in the developed world with a wildlife reservoir of *Mycobacterium bovis* there has been an alarming increase in the incidence of bovine tuberculosis. Data for the year 2000 in Great Britain show a national herd incidence of 2.8%, with an exponential increase in cases in the southwest of England over the past 10 years (www.defra.gov.uk/animalh). The current means of tuberculosis control is the "test and slaughter" strategy, whereby animals giving a positive skin reaction to a crude preparation of mycobacterial antigens are identified as infected and slaughtered. The badger (*Meles meles*) has been suggested to act as a significant source of infection in Great Britain and Ireland, with a large-scale trial currently underway to evaluate the contribution of badger culling to the control of bovine tuberculosis (1). Infection with *M. bovis* has also been described across a range of animals such as buffalo, kudu, lion, and antelope in the Kruger National Park, having severe implications for the biodiversity of this region (2). In New Zealand, the eradication of bovine tuberculosis is confounded by a continuing problem of wildlife reservoirs of *M. bovis*, especially in the brushtail possum (*Trichosurus vulpecula*) (3). The presence of *M. bovis* infection in white-tailed deer in Michigan poses a serious threat to the control and eradication programs for bovine tuberculosis that are in their final stages in the United States (4). There is a clear need for new control strategies if the worldwide threat from bovine tuberculosis is to be eradicated.

The disease is caused by *M. bovis*, a Gram-positive bacillus with zoonotic potential that is highly genetically related to *Mycobacterium tuberculosis*, the causative agent of human tuberculosis (5, 6). Although the human and bovine tubercle bacilli can be differentiated by host range, virulence and physiological features the genetic basis for these differences is unknown. *M. bovis* was also the progenitor of the only current vaccine against tuberculosis, *M. bovis* bacillus Calmette–Guérin, a strain that was attenuated by serial passage of *M. bovis* on potato slices soaked in ox-bile and glycerol over 13 years (7). However, the precise mutations that led to attenuation of bacillus Calmette–Guérin are still unknown, though one key deletion (RD1) appears to have played a role (8).

With the availability of the genome sequence of *M. bovis*, we are now in a position to address the genetic basis of key phenotypic traits of the bovine tubercle bacillus. Here we use comparative analyses to show that deletion of genetic information has been the dominant force in shaping the genome, with *M. bovis* not presenting any unique genes *per se* compared with other members of the *M. tuberculosis* complex. The analyses we present here suggest that variation of cell wall components and gene expression were key to the evolution of *M. bovis*.

Methods

For the shotgun phase, a total of 81,146 reads, or ≈7.7 coverage, was generated from pUC18 and M13mp18 small (1–4 kb) insert libraries by using dye-terminator chemistry on ABI377 or ABI3700 automated DNA sequencers (Applied Biosystems). Assembly of the shotgun data were performed by PHRAP (P. Green, unpublished data). The sequence was finished by using GAP4 as described (5), with an extra 13,000 reads from the pUC libraries performed on ABI3700 machines for finishing purposes. Annotation was managed through the ARTEMIS (www.sanger.ac.uk/Software) tool, with comparisons to public and in-house databases performed by using the BLAST suite and FASTA. Comparative genome analysis was achieved by using the Artemis Comparison Tool (ACT; www.sanger.ac.uk/Software) with single-nucleotide polymorphism (SNP) identification performed by using the EMBOSS package (www.hgmp.mrc.ac.uk/Software/EMBOSS). The sequence and annotation have been deposited in the EMBL database under accession no. BX248333.

Results and Discussion

Genome Features. *M. bovis* AF2122/97 is a fully virulent Great Britain strain isolated in 1997 from a diseased cow suffering

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: SNP, single-nucleotide polymorphism; CDS, coding sequences.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. BX248333).

¶To whom correspondence should be addressed. E-mail: s.v.gordon@vla.defra.gsi.gov.uk.

Table 1. Overview of genome comparison

Feature	*M. bovis* AF2122/97	*M. tuberculosis* H37Rv	*M. tuberculosis* CDC1551
Genome size, bp	4,345,492	4,411,532	4,403,836
G + C, %	65.6	65.6	65.6
Protein coding genes*	3,951	3,995	4,249
Compared to *M. bovis*			
SNPs	–	2,437	2,423
Transitions	–	1,649	1,630
Transversions	–	788	793
Deletions (≥1 bp)	–	205	221
Insertions (≥1 bp)	–	177	245

*The increased number of potential protein coding genes in CDC1551 is caused by the use of a different gene prediction algorithm.

caseous lesions in lung and bronchomediastinal lymph nodes. The genome sequence is 4,345,492 bp in length, arranged in a single circular chromosome with an average G + C content of 65.63% (Table 1). The genome contains 3,952 genes encoding proteins, including a prophage and 42 IS elements (Fig. 1). Strikingly, the genome is >99.95% identical at the nucleotide level to that of *M. tuberculosis*, showing colinearity and no evidence of extensive translocations, duplications or inversions. Before the availability of the *M. bovis* genome sequence, comparative genomics of the *M. tuberculosis* complex had been performed by using hybridization-based methods, exploiting this high degree of sequence identity (9–12). This revealed 11 deletions from the genome of *M. bovis*, ranging in size from ≈1 to 12.7 kb, and these have been confirmed by the sequence data.

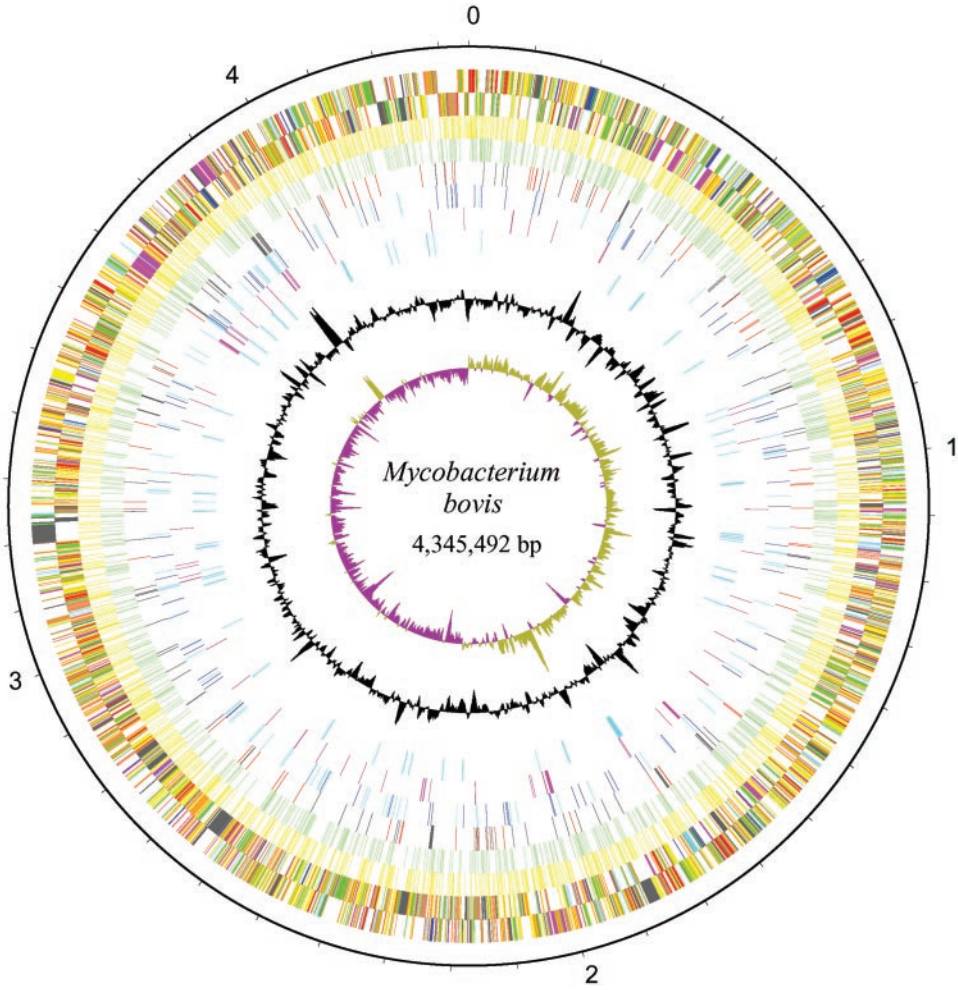

Fig. 1. Circular representation of the *M. bovis* genome. The scale is shown in megabases by the outer black circle. Moving in from the outside, the next two circles show forward and reverse strand CDS, respectively, with colors representing the functional classification. Comparisons with the *M. tuberculosis* H37Rv sequence are then shown, with transitions (yellow) and transversions (green), then insertions (red, 1 bp; black >1 bp) and deletions (dark blue, 1 bp; light blue >1 bp); sequence replacements by novel regions in *M. bovis* are then shown (purple). IS elements and phage (cyan) are displayed in the following circle, with G+C content and then finally GC bias (G-C)/(G+C) shown by using a 20-kb window.

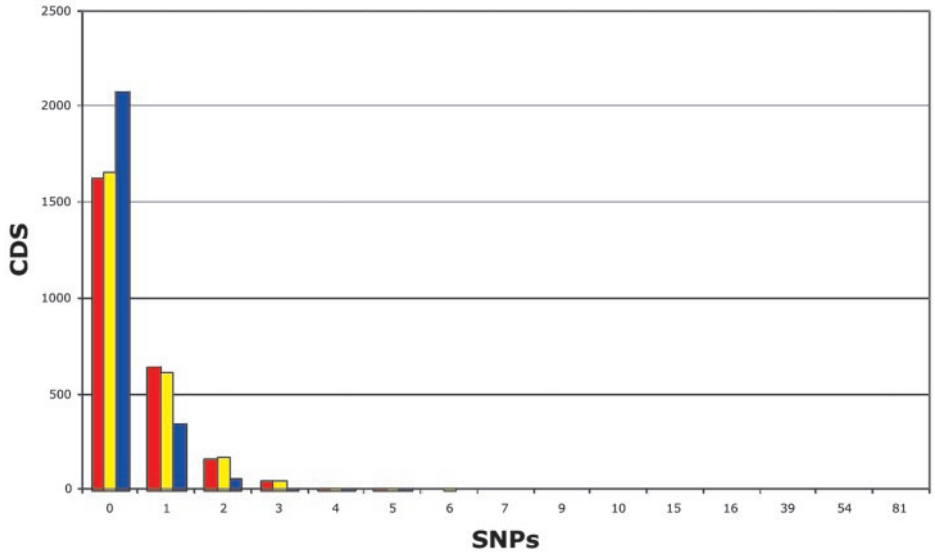

Fig. 2. Tripartite comparison of 2,504 CDS of *M. bovis*, *M. tuberculosis* H37Rv, and *M. tuberculosis* CDC1551. The colors represent the *M. bovis*–*M. tuberculosis* H37Rv comparison (red), *M. bovis*–*M. tuberculosis* CDC1551 (yellow), and *M. tuberculosis* H37Rv–*M. tuberculosis* CDC1551 (blue). The *y* axis shows numbers of CDS, with the *x* axis displaying the numbers of SNPs (both synonymous and nonsynonymous).

Surprisingly, the sequence contains only one locus in *M. bovis*, termed TbD1 (see below), which is absent from the majority of extant *M. tuberculosis* strains. Therefore, at a gross level, deletion has been the dominant mechanism in shaping the *M. bovis* genome.

Comparison with *M. tuberculosis*: SNPs. There are 2,437 SNPs between *M. bovis* and *M. tuberculosis* H37Rv, and 2,423 compared with *M. tuberculosis* CDC1551 (13) (Table 1). SNPs have previously been shown to be responsible for a number of distinctive characteristics of the bovine bacillus. For example, a point mutation in the *pncA* gene in *M. bovis* confers resistance to the key anti-tuberculosis drug pyrazinamide and prevents the accumulation of niacin that is seen in *M. tuberculosis* (14, 15). Direct comparison of 2,504 coding sequences (CDS) of identical length across the three genomes revealed that 1629 and 1656 *M. bovis* CDS are identical in *M. tuberculosis* H37Rv and CDC1551 respectively (Fig. 2). This compares to 2,082 CDS that show no difference between the two *M. tuberculosis* strains. Across these selected CDS, *M. bovis* showed 506 synonymous and 769 nonsynonymous SNPs compared with *M. tuberculosis* H37Rv, with 506 synonymous and 800 nonsynonymous SNPs against *M. tuberculosis* CDC1551. The two *M. tuberculosis* strains showed 339 nonsynonymous and 241 synonymous SNPs, respectively. This analysis not only underlines the conservation of gene sequence across members of the *M. tuberculosis* complex, but also the divergence of *M. bovis* from *M. tuberculosis*. The unexpectedly high frequency of nonsynonymous to synonymous changes may be a product of the close evolutionary relationship between these strains.

Cell Envelope and Antigenic Variation. Cell walls of pathogenic bacteria are known to show variation in protein sequences and macromolecular composition, reflecting selective pressures on these structures. It is therefore notable that the greatest degree of sequence variation between the human and bovine bacilli is found in genes encoding cell wall and secreted proteins (Fig. 3). Variation in genes encoding lipoproteins is seen, with *lppO*, *lpqT*, *lpqG*, and *lprM* deleted or frameshifted, whereas *M. bovis* has a duplicated copy of *lppA*. Similarly, the *M. bovis rpfA* gene, one of a five-membered family encoding secreted proteins that promote the resuscitation of dormant or nongrowing bacilli (16), shows an in-frame deletion of 240 bp that leads to the synthesis of a shorter protein. Whether this affects the function of the protein, or again reflects antigenic variation, is unclear. There is extensive variation in genes encoding the PE-PGRS and PPE protein families (5). Although initially of unknown function, there is now a considerable body of evidence to suggest that at least some of these proteins are surface exposed and play a role in adhesion and immune modulation (17, 18). Between *M. bovis* AF2122/97 and *M. tuberculosis* H37Rv there are blocks of sequence variation in genes encoding 29 different PE-PGRS and 28 PPE proteins resulting from in-frame insertions and deletions, whereas others are frameshifted. Because ≈60% of these proteins differ, this is clearly at odds with the rest of the genome where the majority of genes are identical, and indicates that these gene families can support extensive sequence polymorphism, providing a source of variation for selective pressures to act upon. One of the *M. tuberculosis* PE-PGRS proteins (Rv1759c) binds fibronectin, and this in turn suggests that alterations to the PE-PGRS repertoire might influence host or tissue tropism (19). The *M. bovis* orthologue of Rv1759c is a pseudogene.

A group of known antigens affected by deletions from *M. bovis* is the ESAT-6 family. The ESAT-6 protein was originally described as a potent T cell antigen secreted by *M. tuberculosis* (20), and belongs to a >20-membered family that contains other T cell antigens such as CFP-10 and CFP-7. The demonstration of an interaction between ESAT-6 and CFP-10 suggests that other members of the family may also act in pairs, possibly in a mix-and-match arrangement (21). However, six ESAT-6 proteins, encoded by Rv2346c, Rv2347c, Rv3619c, Rv3620c, Rv3890c (Mb3919c), and Rv3905c (Mb3935c) in *M. tuberculosis*, are missing or altered in *M. bovis* (Fig. 3). The consequences of their loss are difficult to predict, though they may impact on antigen load either singly or in combination.

The most striking degree of variation in the secretome is the elevated expression of two serodominant antigens, MPB70 and MPB83, in the bovine bacillus (22). MPB83 is a glycosylated cell wall-associated protein, whereas MPB70 is a secreted protein that can account for 10% of *M. bovis* culture filtrate proteins (23). Differences are also seen in genes encoding the synthesis (*pks*) and transport (*mmpSL*) of polyketides and complex lipids with polyketide moieties (Fig. 3). These lipids are major factors

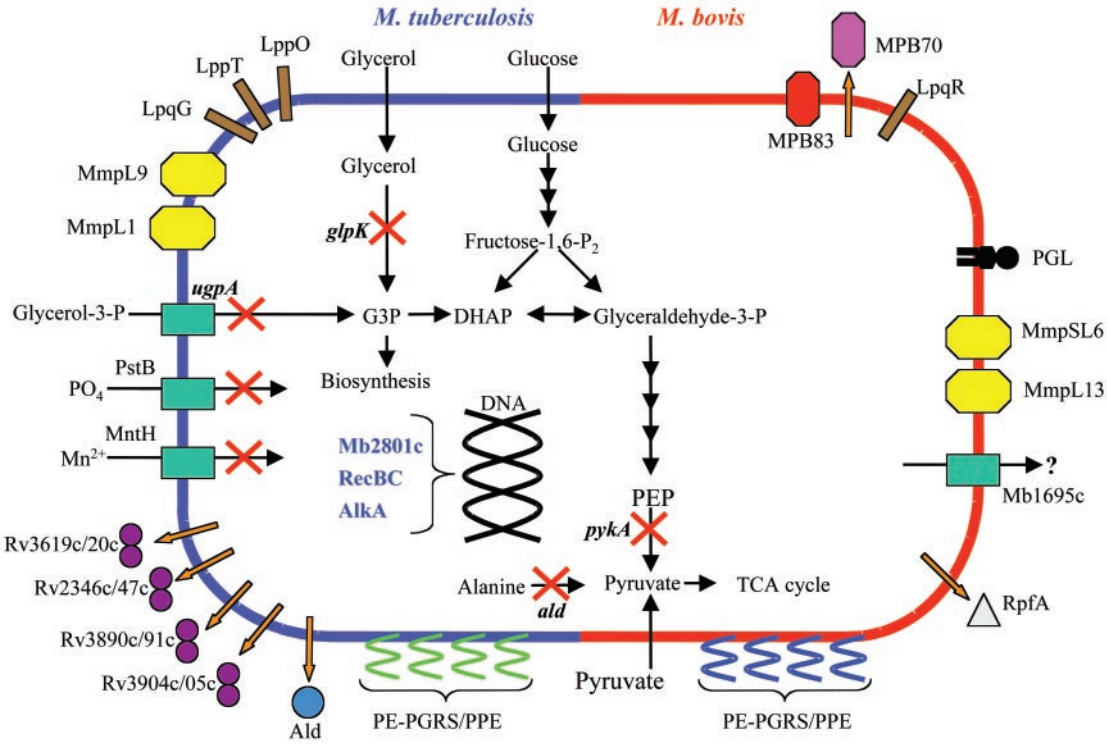

Fig. 3. Schematic of the major differences between *M. bovis* AF2122/97 and *M. tuberculosis* H37Rv. The blue and red lines represent the cell wall, with blue showing *M. tuberculosis* and red showing *M. bovis*. Surface-exposed and transport molecules particular to each bacillus are shown embedded in the wall. Because the large number of differences in the PE-PGRS and PPE are beyond the scope of this diagram, they are merely represented by surface-exposed molecules. Differentially secreted proteins (orange arrows) are shown in each half. The interior of the diagram shows the key steps in carbohydrate metabolism, with the red crosses showing where lesions occur in *M. bovis*. Proteins that interact with DNA which are inactivated in *M. bovis* are shown in blue. PGL, phenolic glycolipid; G3P, glycerol 3-phosphate; DHAP, dihydroxyacetone phosphate; PEP, phosphoenolpyruvate; Ald, alanine dehydrogenase.

in inducing host pathologies that create more favorable environments for the pathogens (24, 25). The genes *pks1*, *mmpL13*, and Mb1695c (a putative macrolide transporter adjacent to the *pks10/7/8/17/9/11* cluster) could be translated to functional products in *M. bovis*, but are disrupted in *M. tuberculosis*. The opposite is the case (i.e., disrupted in *M. bovis*) for the linked *pks6* and *mmpL1* genes and *mmpL9*. It has been shown functionally that *pks1* codes for the biosynthesis of the major phenolic glycolipid of *M. bovis* and *Mycobacterium canettii*, whereas in strains where *pks1* is disrupted, such as *M. tuberculosis*, no such lipid is produced (26). It has been suggested that many *pks* gene products that have never been seen in axenic culture may only be produced by tubercle bacilli in the host (5). Thus, it is curious that one of them, *pks6*, is disrupted in *M. bovis*, because inactivation of this gene has been shown to attenuate *M. tuberculosis* in the mouse model (24).

The TbD1 locus, containing the gene *mmpS6* and the 5′ region of *mmpL6*, is absent from a majority of *M. tuberculosis* strains (27). Deletion of TbD1 may therefore prevent trafficking of specific lipids to the cell wall of *M. tuberculosis*. Furthermore, a deletion of 808 bp is proximal to the TbD1 region and truncates the *treY* gene. As *treY* encodes a maltooligosyltrehalose synthase, an enzyme in a pathway for trehalose production (two other pathways are intact) (28), its deletion in *M. bovis* may have an effect on the range of trehalose-based glycolipids that are produced. Deletion analysis has revealed that the *treY* lesion is not present in all strains of *M. bovis*, suggesting utility as a marker for deep phylogeny. Disruption of the Rv1373 orthologue (Mb1407/8) accounts for the lack of sulfated-lipids in the envelope, because the encoded enzyme has been shown functionally to be a glycolipid sulfotransferase (29). Overall, these differences could have major effects on phenotype and host interaction.

Global Gene Regulation. The *M. bovis* and *M. tuberculosis* genomes are >99.95% identical at the nucleotide level. However, an amplification of difference is achieved when changes are in regulatory genes as perforce each one affects the expression of a wide range of genes. In fact, many differences would appear to inactivate genes encoding regulatory proteins in *M. bovis*. The *M. tuberculosis alkA* gene codes for a DNA repair protein with an N-terminal regulatory domain (activated by DNA damage) and a C-terminal DNA glycosylase (30). However the *M. bovis alkA* contains a frameshift at the start of the CDS, leading to the synthesis of a truncated protein. Based on the *Escherichia coli* model of AlkA function, it is possible that this lesion impairs the ability of *M. bovis* to respond to nitrosative stress and induce an effective DNA repair response. An AsnC/Lrp family regulator encoded by Mb2801c (Rv2779c), which has been shown to be up-regulated in response to nutrient starvation in *M. tuberculosis* (31), shows an 8-aa deletion in the core of the protein that may affect tertiary structure or DNA binding. The *pknH* gene encoding a serine/threonine protein kinase shows an internal deletion and sequence variation relative to the *M. tuberculosis* orthologue. Although the putative active sites are conserved, this variation may affect substrate specificity. Another serine/threonine kinase gene, *pknD*, is a pseudogene in *M. bovis* (32), with adjacent gene clusters also showing disruptions, including two *pst* clusters where the sole *pstB* orthologue is frameshifted. This would be expected to prevent high affinity phosphate uptake, although the *phoT* gene, which encodes a protein with similarity to PstB, may complement this activity. In the same region a frameshift leads to fusion of the *mntH*-encoded manganese transporter with the preceding CDS leading to a 287-aa N-terminal hydrophilic extension, possibly preventing correct positioning of the MntH transporter in the membrane. Inacti-

vation of *mntH* does not affect the virulence of *M. tuberculosis* (33). However, because Mn^{2+} ions are required for regulatory functions such as relaxation of the stringent response (34), lesions in phosphate and manganese transport could affect global gene regulation in *M. bovis*. It is also probable that the ability to reduce nitrate, one of the characteristic tests that differentiates human and bovine tubercle bacilli, is linked to gene regulation. Classically, *M. bovis* is described as being nitrate reductase negative (35). However, Bange and colleagues (36) have shown that growth of *M. bovis* bacillus Calmette–Guérin under oxygen-limiting conditions leads to expression of nitrate reductase activity. Variation in expression networks is undoubtedly central to many phenotypic differences between the tubercle bacilli.

Insights on *in Vivo* Growth. One of the key *in vitro* differences between *M. bovis* and *M. tuberculosis* is a requirement for pyruvate when glycerol is the sole carbon source (35). This presumably reflects a defect in the metabolism of glycerol by *M. bovis*. It is therefore intriguing that *M. bovis* presents multiple lesions in carbohydrate catabolism. The *glpK* gene of *M. bovis* AF2122/97, encoding glycerol kinase, is a pseudogene, preventing the phosphorylation of glycerol and therefore its use as a carbon source. Furthermore, *ugpA*, encoding part of the putative ATP-binding cassette transporter for glycerol-3-phosphate, is also a pseudogene in *M. bovis*. In addition, a frameshift that fuses the genes encoding the iron–sulfur protein (*frdB*) and one of the membrane-spanning domains (*frdC*) of fumarate reductase could affect positioning of this key anaerobic enzyme in the membrane. Strikingly, we have found that *M. bovis* lacks pyruvate kinase activity, with *pykA* containing a point mutation that would affect binding of the Mg^{2+} cofactor. Pyruvate kinase catalyses the final irreversible step in glycolysis, the dephosphorylation of phosphoenolpyruvate to pyruvate. Hence, in *M. bovis*, glycolytic intermediates are blocked from feeding into oxidative metabolism. Moreover, in another reaction leading to pyruvate the *ald* gene, encoding alanine dehydrogenase, is a pseudogene, therefore blocking the conversion of alanine to pyruvate.

Our initial analysis has shown that although the frameshift in the *glpK* of *M. bovis* AF2122/97 is not universally present in *M. bovis* strains, other SNPs in genes of carbohydrate catabolism were identical in the *M. bovis* strains tested. Also, the *glpK*, *ugpA* and *pykA* mutations are not present in the vaccine *M. bovis* bacillus Calmette–Guérin Pasteur, a strain that does not require pyruvate to be added to glycerinated media and that possesses glycerol kinase and pyruvate kinase activity (unpublished observations). The creation of *M. bovis* bacillus Calmette–Guérin by the serial passage of a strain of *M. bovis* for 13 years on glycerol-soaked potato slices must therefore have selected for the correction of key lesions in carbohydrate metabolism (37). It remains to be seen whether alterations in metabolism played a role in the attenuation of *M. bovis* bacillus Calmette–Guérin. However, it is clear that *in vivo M. bovis* must rely on amino acids or fatty acids as a carbon source for energy metabolism.

Genome Downsizing. Deletion of information is the dominant trend in the *M. bovis* genome. This has parallels with the genome of *Mycobacterium leprae*, which has lost >1.1 Mb and accumulated >1,100 pseudogenes during reductive evolution (38). Indeed, many of the genes either deleted or inactivated are common in the two organisms. For example, genes involved in transport and cell surface structures (*pstB*, *ugpA*, *mce3A-F*, *lppO*, *lpqG*, *lprM*, *pks6*, *mmpL1*, *mmpL9*, Rv1510, Rv1508, Rv1371), fatty acid metabolism (*fadE22*, *echA1*), cofactor biosynthesis (*moaE*, *moaC2*), detoxification (*ephA*, *ephF*, *alkA*), and intermediary metabolism (*epiA*, *gmdA*) are pseudogenes or deleted in both bacilli. Similarly, *M. leprae* and *M. bovis* have lost the AtsA system for recycling sulfate (39). AtsA is an arylsulphatase that catalyses the hydrolysis of sulfate esters to release inorganic sulfate. Loss of this function may reflect the lack of sulfated glycolipid in these two mycobacteria. This builds on work that showed that *M. bovis* bacillus Calmette–Guérin does not need sulfate *in vivo* as a *cysA* mutant, inactivated in the sole transporter for sulfate, persisted *in vivo* as well as the parent strain (40). It also reflects the situation in *M. leprae*, where *cysTWA* are pseudogenes. Furthermore, *recBCD* are deleted in *M. leprae*, whereas *recB* is frameshifted in *M. bovis*. This frameshift removes the C-terminal domain of RecB, which is essential for the nuclease activity of RecBCD (41). However, as *M. bovis* can support homologous recombination it is likely that polar effects on RecD act to suppress defects in recombination (42).

Conclusions

It has long been thought that human tuberculosis had its origin as a zoonosis, with *M. bovis* jumping the species barrier and host adapting to humans to become *M. tuberculosis* at the time of cattle domestication 10,000–15,000 years ago. However, using deletion analysis a new scenario for the evolution of the *M. tuberculosis* complex has recently been proposed that places *M. tuberculosis* closer to the common progenitor of the complex than *M. bovis* (27). The completion of the *M. bovis* genome sequence has confirmed the predictions of this new scenario, showing that *M. bovis* has evolved from a progenitor of the *M. tuberculosis* complex as a clone showing distinct host preference.

The possibility exists that deletion events from the genome of *M. bovis* represent "black holes," i.e., the loss of genes that are detrimental to the pathogenic lifestyle in a specific niche (43). However, the analysis we present here suggests that although the adaptation process did not rely on the presence of specific virulence genes *per se*, alterations in gene expression and exposed components of the cell envelope played leading roles. The genome sequence will therefore have a major impact on our understanding of the evolution, host adaptation and pathobiology of tuberculosis and, in the longer term, on the generation of vaccine candidates and diagnostic reagents to combat disease.

This article is dedicated to the memory of Jean-Christophe Camus. We thank Noel Smith, John Maynard Smith, Roland Brosch, Christiane Bouchier, and Rik Myers for advice and discussion. This work was funded by the Department for Environment, Food and Rural Affairs (Great Britain), The Wellcome Trust, the Association Française Raoul Follereau, the Génopole Program, and the Institut Pasteur.

1. Krebs, J. (1997) *Bovine Tuberculosis in Cattle and Badgers: Report to the Rt. Hon. Dr. Jack Cunningham MP by the Independent Scientific Review Group* (Department for Environment, Food and Rural Affairs, London).
2. Weyer, K., Fourie, P. B., Durrheim, D., Lancaster, J., Haslov, K. & Bryden, H. (1999) *Int. J. Tuberc. Lung Dis.* **3,** 1113–1119.
3. Morris, R. & Pfeiffer, D. (1995) *N. Zealand Vet. J.* **43,** 256–265.
4. Payeur, J. B., Church, S., Mosher, L., Robinson-Dunn, B., Schmitt, S. & Whipple, D. (2002) *Ann. N.Y. Acad. Sci.* **969,** 259–261.
5. Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., 3rd, *et al.* (1998) *Nature* **393,** 537–544.
6. O'Reilly, L. M. & Daborn, C. J. (1995) *Tubercle Lung Dis.* **76,** 1–46.
7. Calmette, A. (1927) *La Vaccination Preventive Contre la Tuberculose* (Masson et Cie, Paris).
8. Pym, A. S., Brodin, P., Brosch, R., Huerre, M. & Cole, S. T. (2002) *Mol. Microbiol.* **46,** 709–717.
9. Rauzier, J., Gormley, E., Gutierrez, M. C., Kassa-Kelembho, E., Sandall, L. J., Dupont, C., Gicquel, B. & Murray, A. (1999) *Microbiology* **145,** 1695–1701.
10. Gordon, S. V., Brosch, R., Billault, A., Garnier, T., Eiglmeier, K. & Cole, S. T. (1999) *Mol. Microbiol.* **32,** 643–655.
11. Behr, M. A., Wilson, M. A., Gill, W. P., Salamon, H., Schoolnik, G. K., Rane, S. & Small, P. M. (1999) *Science* **284,** 1520–1523.

12. Zumarraga, M., Bigi, F., Alito, A., Romano, M. I. & Cataldi, A. (1999) *Microbiology* **145,** 893–897.
13. Fleischmann, R. D., Alland, D., Eisen, J. A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., *et al.* (2002) *J. Bacteriol.* **184,** 5479–5490.
14. Boshoff, H. I., Mizrahi, V. & Barry, C. E., 3rd (2002) *J. Bacteriol.* **184,** 2167–2172.
15. Scorpio, A. & Zhang, Y. (1996) *Nat. Med.* **2,** 662–667.
16. Mukamolova, G. V., Turapov, O. A., Young, D. I., Kaprelyants, A. S., Kell, D. B. & Young, M. (2002) *Mol. Microbiol.* **46,** 623–635.
17. Banu, S., Honore, N., Saint-Joanis, B., Philpott, D., Prevost, M. C. & Cole, S. T. (2002) *Mol. Microbiol.* **44,** 9–19.
18. Brennan, M. J., Delogu, G., Chen, Y., Bardarov, S., Kriakov, J., Alavi, M. & Jacobs, W. R., Jr. (2001) *Infect. Immun.* **69,** 7326–7333.
19. Espitia, C., Laclette, J. P., Mondragon-Palomino, M., Amador, A., Campuzano, J., Martens, A., Singh, M., Cicero, R., Zhang, Y. & Moreno, C. (1999) *Microbiology* **145,** 3487–3495.
20. Sorensen, A. L., Nagai, S., Houen, G., Andersen, P. & Andersen, A. B. (1995) *Infect. Immun.* **63,** 1710–1717.
21. Renshaw, P. S., Panagiotidou, P., Whelan, A., Gordon, S. V., Hewinson, R. G., Williamson, R. A. & Carr, M. D. (2002) *J. Biol. Chem.* **277,** 21598–21603.
22. Hewinson, R. G., Michell, S. L., Russell, W. P., McAdam, R. A. & Jacobs, W. R., Jr. (1996) *Scand. J. Immunol.* **43,** 490–499.
23. Nagai, S., Matsumoto, J. & Nagasuga, T. (1981) *Infect. Immun.* **31,** 1152–1160.
24. Camacho, L. R., Ensergueix, D., Perez, E., Gicquel, B. & Guilhot, C. (1999) *Mol. Microbiol.* **34,** 257–267.
25. Cox, J. S., Chen, B., McNeil, M. & Jacobs, W. R., Jr. (1999) *Nature* **402,** 79–83.
26. Constant, P., Perez, E., Malaga, W., Laneelle, M. A., Saurel, O., Daffe, M. & Guilhot, C. (2002) *J. Biol. Chem.* **277,** 38148–38158.
27. Brosch, R., Gordon, S. V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., Garnier, T., Gutierrez, C., Hewinson, G., Kremer, K., *et al.* (2002) *Proc. Natl. Acad. Sci. USA* **99,** 3684–3689.
28. De Smet, K. A., Weston, A., Brown, I. N., Young, D. B. & Robertson, B. D. (2000) *Microbiology* **146,** 199–208.
29. Rivera-Marrero, C. A., Ritzenthaler, J. D., Newburn, S. A., Roman, J. & Cummings, R. D. (2002) *Microbiology* **148,** 783–792.
30. Mizrahi, V. & Andersen, S. J. (1998) *Mol. Microbiol.* **29,** 1331–1339.
31. Betts, J. C., Lukey, P. T., Robb, L. C., McAdam, R. A. & Duncan, K. (2002) *Mol. Microbiol.* **43,** 717–731.
32. Peirs, P., Parmentier, B., De Wit, L. & Content, J. (2000) *FEMS Microbiol. Lett.* **188,** 135–139.
33. Domenech, P., Pym, A. S., Cellier, M., Barry, C. E., 3rd, & Cole, S. T. (2002) *FEMS Microbiol. Lett.* **207,** 81–86.
34. Avarbock, D., Avarbock, A. & Rubin, H. (2000) *Biochemistry* **39,** 11640–11648.
35. Wayne, L. G. (1984) in *The Mycobacteria: A Sourcebook*, ed. Wayne, L. G. (Dekker, New York), Vol. A, pp. 25–65.
36. Weber, I., Fritz, C., Ruttkowski, S., Kreft, A. & Bange, F. C. (2000) *Mol. Microbiol.* **35,** 1017–1025.
37. Calmette, A. & Guérin, C. (1909) *C. R. Acad. Sci. Paris* **149,** 716–718.
38. Cole, S. T., Eiglmeier, K., Parkhill, J., James, K. D., Thomson, N. R., Wheeler, P. R., Honore, N., Garnier, T., Churcher, C., Harris, D., *et al.* (2001) *Nature* **409,** 1007–1011.
39. Hummerjohann, J., Laudenbach, S., Retey, J., Leisinger, T. & Kertesz, M. A. (2000) *J. Bacteriol.* **182,** 2055–2058.
40. Wooff, E., Michell, S. L., Gordon, S. V., Chambers, M. A., Bardarov, S., Jacobs, W. R., Hewinson, R. G. & Wheeler, P. R. (2002) *Mol. Microbiol.* **43,** 653–663.
41. Jockovich, M. E. & Myers, R. S. (2001) *Mol. Microbiol.* **41,** 949–962.
42. Amundsen, S. K., Taylor, A. F. & Smith, G. R. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 7399–7404.
43. Maurelli, A. T., Fernandez, R. E., Bloch, C. A., Rode, C. K. & Fasano, A. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 3943–3948.